

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

## Discrete Applied Mathematics

journal homepage: [www.elsevier.com/locate/dam](http://www.elsevier.com/locate/dam)Generalized roof duality and bisubmodular functions<sup>☆</sup>Vladimir Kolmogorov<sup>\*</sup>

Institute of Science and Technology (IST), Austria

## ARTICLE INFO

## Article history:

Received 20 December 2010

Received in revised form 10 October 2011

Accepted 21 October 2011

Available online 25 November 2011

## Keywords:

Pseudo-boolean optimization

Roof duality

Bisubmodularity

## ABSTRACT

Consider a convex relaxation  $\hat{f}$  of a pseudo-Boolean function  $f$ . We say that the relaxation is *totally half-integral* if  $\hat{f}(\mathbf{x})$  is a polyhedral function with half-integral extreme points  $\mathbf{x}$ , and this property is preserved after adding an arbitrary combination of constraints of the form  $x_i = x_j$ ,  $x_i = 1 - x_j$ , and  $x_i = \gamma$  where  $\gamma \in \{0, 1, \frac{1}{2}\}$  is a constant. A well-known example is the *roof duality* relaxation for quadratic pseudo-Boolean functions  $f$ . We argue that total half-integrality is a natural requirement for generalizations of roof duality to arbitrary pseudo-Boolean functions.

Our contributions are as follows. First, we provide a complete characterization of totally half-integral relaxations  $\hat{f}$  by establishing a one-to-one correspondence with *bisubmodular functions*. Second, we give a new characterization of bisubmodular functions. Finally, we show some relationships between general totally half-integral relaxations and relaxations based on the roof duality.

On the conceptual level, our results show that bisubmodular functions provide a natural generalization of the roof duality approach to higher-order terms. This can be viewed as a non-submodular analogue of the fact that submodular functions generalize the  $s$ - $t$  minimum cut problem with non-negative weights to higher-order terms.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Let  $V$  be a set of  $|V| = n$  nodes and  $\mathcal{B} \subset \mathcal{K}^{1/2} \subset \mathcal{K}$  be the following sets:

$$\mathcal{B} = \{0, 1\}^V \quad \mathcal{K}^{1/2} = \left\{0, \frac{1}{2}, 1\right\}^V \quad \mathcal{K} = [0, 1]^V.$$

A function  $f : \mathcal{B} \rightarrow \mathbb{R}$  is called *pseudo-Boolean*. In this paper we consider convex relaxations  $\hat{f} : \mathcal{K} \rightarrow \mathbb{R}$  of  $f$  which we call *totally half-integral*:

**Definition 1.** (a) Function  $\hat{f} : \mathcal{P} \rightarrow \mathbb{R}$  where  $\mathcal{P} \subseteq \mathcal{K}$  is called *half-integral* if it is a convex polyhedral function such that all extreme points of the epigraph  $\{(\mathbf{x}, z) \mid \mathbf{x} \in \mathcal{P}, z \geq \hat{f}(\mathbf{x})\}$  have the form  $(\mathbf{x}, \hat{f}(\mathbf{x}))$  where  $\mathbf{x} \in \mathcal{K}^{1/2}$ . (b) Function  $\hat{f} : \mathcal{K} \rightarrow \mathbb{R}$  is called *totally half-integral* if restrictions  $\hat{f}|_{\mathcal{P}} : \mathcal{P} \rightarrow \mathbb{R}$  are half-integral for all subsets  $\mathcal{P} \subseteq \mathcal{K}$  obtained from  $\mathcal{K}$  by adding an arbitrary combination of constraints of the form  $x_i = x_j$ ,  $x_i = \bar{x}_j$ , and  $x_i = \gamma$  for points  $\mathbf{x} \in \mathcal{K}$ . Here  $i, j$  denote nodes in  $V$ ,  $\gamma$  denotes a constant in  $\{0, 1, \frac{1}{2}\}$ , and  $\bar{z} \equiv 1 - z$ .

<sup>☆</sup> A preliminary version of this paper appeared in the proceedings of Neural Information Processing Systems conference (NIPS), 2010.

<sup>\*</sup> Fax: +43 0 2243 9000 2000.

E-mail address: [vnk@ist.ac.at](mailto:vnk@ist.ac.at).

A well-known example of a totally half-integral relaxation is the *roof duality relaxation* for quadratic pseudo-Boolean functions  $f(\mathbf{x}) = \sum_i c_i x_i + \sum_{(i,j)} c_{ij} x_i x_j$  studied by Hammer et al. [13]. It is known to possess the *persistence* property: for any half-integral minimizer  $\hat{\mathbf{x}} \in \arg \min \hat{f}(\hat{\mathbf{x}})$  there exists minimizer  $\mathbf{x} \in \arg \min f(\mathbf{x})$  such that  $x_i = \hat{x}_i$  for all nodes  $i$  with integral component  $\hat{x}_i$ . This property is quite important in practice as it allows to reduce the size of the minimization problem when  $\hat{\mathbf{x}} \neq \frac{1}{2}$ . The set of nodes with guaranteed optimal solution can sometimes be increased further using the PROBE technique [6], which also relies on persistency.

The goal of this paper is to generalize the roof duality approach to arbitrary pseudo-Boolean functions. The total half-integrality is a very natural requirement of such generalizations, as discussed later in this section. As we prove, total half-integrality implies persistency.

We provide a complete characterization of totally half-integral relaxations. Namely, we prove in Section 2 that if  $\hat{f} : \mathcal{K} \rightarrow \mathbb{R}$  is totally half-integral then its restriction to  $\mathcal{K}^{1/2}$  is a *bisubmodular function*, and conversely any bisubmodular function can be extended to a totally half-integral relaxation.

**Definition 2.** Function  $f : \mathcal{K}^{1/2} \rightarrow \mathbb{R}$  is called *bisubmodular* if

$$f(\mathbf{x} \sqcap \mathbf{y}) + f(\mathbf{x} \sqcup \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{K}^{1/2} \quad (1)$$

where binary operators  $\sqcap, \sqcup : \mathcal{K}^{1/2} \times \mathcal{K}^{1/2} \rightarrow \mathcal{K}^{1/2}$  are defined component-wise as follows:

$$\begin{array}{c|ccc} \sqcap & 0 & \frac{1}{2} & 1 \\ \hline 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} & 1 \end{array} \quad \begin{array}{c|ccc} \sqcup & 0 & \frac{1}{2} & 1 \\ \hline 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 \\ 1 & \frac{1}{2} & 1 & 1 \end{array} \quad (2)$$

As our second contribution, we give a new characterization of bisubmodular functions (Section 3). Using this characterization, we then prove several results showing links with the roof duality relaxation (Section 4).

### 1.1. Applications

This work has been motivated by computer vision applications. A fundamental task in vision is to infer pixel properties from observed data. These properties can be the type of object to which the pixel belongs, distance to the camera, pixel intensity before being corrupted by noise, etc. The popular *MAP-MRF approach* casts the inference task as an *energy minimization* problem with the objective function of the form  $f(\mathbf{x}) = \sum_C f_C(\mathbf{x})$  where  $C \subset V$  are subsets of neighboring pixels of small cardinality ( $|C| = 1, 2, 3, \dots$ ) and terms  $f_C(\mathbf{x})$  depend only on labels of pixels in  $C$ .

For some vision applications the roof duality approach [13] has shown a good performance [29,31,22,23,32,1,16,17].<sup>1</sup> Functions with higher-order terms are steadily gaining popularity in computer vision [30,32,1,16,17]; it is generally accepted that they correspond to better image models. Therefore, studying generalizations of roof duality to arbitrary pseudo-Boolean functions is an important task. In such generalizations the *total* half-integrality property is essential. Indeed, in practice, the relaxation  $\hat{f}$  is obtained as the sum of relaxations  $\hat{f}_C$  constructed for each term independently. Some of these terms can be  $c|x_i - x_j|$  and  $c|x_i + x_j - 1|$ . If  $c$  is sufficiently large, then applying the roof duality relaxation to these terms would yield constraints  $x_i = x_j$  and  $x = \bar{x}_j$  present in the definition of total half-integrality. Constraints  $x_i = \gamma \in \{0, 1, \frac{1}{2}\}$  can also be simulated via the roof duality, e.g.  $x_i = x_j, x_i = \bar{x}_j$  for the same pair of nodes  $i, j$  implies  $x_i = x_j = \frac{1}{2}$ .

### 1.2. Related work

**Half-integrality.** There is a vast literature on using half-integral relaxations for various combinatorial optimization problems. In many cases these relaxations lead to 2-approximation algorithms. Below we list a few representative papers.

The earliest work recognizing half-integrality of polytopes with certain pairwise constraints was perhaps by Balinski [3], while the persistency property goes back to Nemhauser and Trotter [27] who considered the vertex cover problem. Hammer et al. [13] established that these properties hold for the *roof duality relaxation* for quadratic pseudo-Boolean functions. Their work was generalized to arbitrary pseudo-Boolean functions by Lu and Williams [24]. (The relaxation in [24] relied on converting function  $f$  to a multinomial representation; see Section 4 for more details.) Hochbaum [14,15] gave a class of integer problems with half-integral relaxations. Very recently, Iwata and Nagano [18] formulated a half-integral relaxation for the problem of minimizing submodular function  $f(\mathbf{x})$  under constraints of the form  $x_i + x_j \geq 1$ .

<sup>1</sup> In many vision problems variables  $x_i$  are not Boolean. However, such problems are often reduced to a sequence of minimization problems with *Boolean* variables using iterative move-making algorithms, e.g. using *expansion moves* [9] or *fusion moves* [22,23,32,17].

In computer vision, several researchers considered the following scheme: given a function  $f(\mathbf{x}) = \sum f_C(\mathbf{x})$ , convert terms  $f_C(\mathbf{x})$  to *quadratic* pseudo-Boolean functions by introducing auxiliary Boolean variables, and then apply the roof duality relaxation to the latter. Woodford et al. [32] used this technique for the stereo reconstruction problem, while Ali et al. [1] and Ishikawa [16] explored different conversions to quadratic functions.

To the best of our knowledge, all examples of totally half-integral relaxations proposed so far belong to the class of *submodular relaxations*, which is defined in Section 4. They form a subclass of more general bisubmodular relaxations.

**Bisubmodularity.** Bisubmodular functions were introduced by Chandrasekaran and Kadiyali as rank functions of (*poly*-)*pseudomatroids* [10,19]. Independently, Bouchet [7] introduced the concept of  $\Delta$ -matroids which is equivalent to pseudomatroids. Bisubmodular functions and their generalizations have also been considered by Qi [28], Nakamura [26], Bouchet and Cunningham [8] and Fujishige [11]. The notion of the *Lovász extension* of a bisubmodular function introduced by Qi [28] will be of particular importance for our work (see next section).

It has been shown that some submodular minimization algorithms can be generalized to bisubmodular functions. Qi [28] showed the applicability of the ellipsoid method. A weakly polynomial combinatorial algorithm for minimizing bisubmodular functions was given by Fujishige and Iwata [12], and a strongly polynomial version was given by McCormick and Fujishige [25].

Recently, we introduced *strongly* and *weakly tree-submodular* functions [21] that generalize bisubmodular functions.

## 2. Total half-integrality and bisubmodularity

The first result of this paper is the following theorem.

**Theorem 3.** *If  $\hat{f} : \mathcal{K} \rightarrow \mathbb{R}$  is a totally half-integral relaxation then its restriction to  $\mathcal{K}^{1/2}$  is bisubmodular. Conversely, if function  $f : \mathcal{K}^{1/2} \rightarrow \mathbb{R}$  is bisubmodular then it has a unique totally half-integral extension  $\hat{f} : \mathcal{K} \rightarrow \mathbb{R}$ .*

This section is devoted to the proof of Theorem 3. Denote  $\mathcal{L} = [-1, 1]^V$ ,  $\mathcal{L}^{1/2} = \{-1, 0, 1\}^V$ . It will be convenient to work with functions  $\hat{h} : \mathcal{L} \rightarrow \mathbb{R}$  and  $h : \mathcal{L}^{1/2} \rightarrow \mathbb{R}$  obtained from  $\hat{f}$  and  $f$  via a linear change of coordinates  $x_i \mapsto 2x_i - 1$ . Under this change totally half-integral relaxations are transformed to *totally integral* relaxations:

**Definition 4.** Let  $\hat{h} : \mathcal{L} \rightarrow \mathbb{R}$  be a function of  $n$  variables. (a)  $\hat{h}$  is called *integral* if it is a convex polyhedral function such that all extreme points of the epigraph  $\{(\mathbf{x}, z) \mid \mathbf{x} \in \mathcal{L}, z \geq \hat{h}(\mathbf{x})\}$  have the form  $(\mathbf{x}, \hat{h}(\mathbf{x}))$  where  $\mathbf{x} \in \mathcal{L}^{1/2}$ . (b)  $\hat{h}$  is called *totally integral* if it is integral and for an arbitrary ordering of nodes the following functions of  $n - 1$  variables (if  $n > 1$ ) are totally integral:

$$\begin{aligned}\hat{h}'(x_1, \dots, x_{n-1}) &= \hat{h}(x_1, \dots, x_{n-1}, x_{n-1}) \\ \hat{h}'(x_1, \dots, x_{n-1}) &= \hat{h}(x_1, \dots, x_{n-1}, -x_{n-1}) \\ \hat{h}'(x_1, \dots, x_{n-1}) &= \hat{h}(x_1, \dots, x_{n-1}, \gamma) \quad \text{for any constant } \gamma \in \{-1, 0, 1\}.\end{aligned}$$

The definition of a bisubmodular function is adapted as follows: function  $h : \mathcal{L}^{1/2} \rightarrow \mathbb{R}$  is bisubmodular if inequality (1) holds for all  $\mathbf{x}, \mathbf{y} \in \mathcal{L}^{1/2}$  where operations  $\sqcap, \sqcup$  are defined by tables (2) after replacements  $0 \mapsto -1, \frac{1}{2} \mapsto 0, 1 \mapsto 1$ . To prove Theorem 3, it suffices to establish a link between totally integral relaxations  $\hat{h} : \mathcal{L} \rightarrow \mathbb{R}$  and bisubmodular functions  $h : \mathcal{L}^{1/2} \rightarrow \mathbb{R}$ . We can assume without loss of generality that  $\hat{h}(\mathbf{0}) = h(\mathbf{0}) = 0$ , since adding a constant to the functions does not affect the theorem.

A pair  $\omega = (\pi, \sigma)$  where  $\pi : V \rightarrow \{1, \dots, n\}$  is a permutation of  $V$  and  $\sigma \in \{-1, 1\}^V$  will be called a *signed ordering*. Let us rename nodes in  $V$  so that  $\pi(i) = i$ . To each signed ordering  $\omega$  we associate labelings  $\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n \in \mathcal{L}^{1/2}$  as follows:

$$\mathbf{x}^0 = (0, 0, \dots, 0) \quad \mathbf{x}^1 = (\sigma_1, 0, \dots, 0) \quad \dots \quad \mathbf{x}^n = (\sigma_1, \sigma_2, \dots, \sigma_n) \quad (3)$$

where nodes are ordered according to  $\pi$ .

Consider function  $h : \mathcal{L}^{1/2} \rightarrow \mathbb{R}$  with  $h(\mathbf{0}) = 0$ . Its *Lovász extension*  $\hat{h} : \mathbb{R}^V \rightarrow \mathbb{R}$  is defined in the following way [28]. Given a vector  $\mathbf{x} \in \mathbb{R}^V$ , select a signed ordering  $\omega = (\pi, \sigma)$  as follows: (i) choose  $\pi$  so that values  $|x_i|$ ,  $i \in V$  are non-increasing, and rename nodes accordingly so that  $|x_1| \geq \dots \geq |x_n|$ ; (ii) if  $x_i \neq 0$  set  $\sigma_i = \text{sign}(x_i)$ , otherwise choose  $\sigma_i \in \{-1, 1\}$  arbitrarily. It is not difficult to check that

$$\mathbf{x} = \sum_{i=1}^n \lambda_i \mathbf{x}^i \quad (4a)$$

where labelings  $\mathbf{x}^i$  are defined in (3) (with respect to the selected signed ordering) and  $\lambda_i = |x_i| - |x_{i+1}|$  for  $i = 1, \dots, n-1$ ,  $\lambda_n = |x_n|$ . The value of the Lovász extension is now defined as

$$\hat{h}(\mathbf{x}) = \sum_{i=1}^n \lambda_i h(\mathbf{x}^i). \quad (4b)$$

**Theorem 5** ([28]). Function  $h$  is bisubmodular if and only if its Lovász extension  $\hat{h}$  is convex on  $\mathcal{L}$ .<sup>2</sup>

Let  $\mathcal{L}_\omega$  be the set of vectors in  $\mathcal{L}$  for which signed ordering  $\omega = (\pi, \sigma)$  can be selected. Clearly,  $\mathcal{L}_\omega = \{\mathbf{x} \in \mathcal{L} \mid |x_1| \geq \dots \geq |x_n|, x_i \sigma_i \geq 0 \forall i \in V\}$ . It is easy to check that  $\mathcal{L}_\omega$  is the convex hull of  $n + 1$  points (3). Eqs. (4) imply that  $\hat{h}$  is linear on  $\mathcal{L}_\omega$  and coincides with  $h$  in each corner  $\mathbf{x}^0, \dots, \mathbf{x}^n$ .

**Lemma 6.** Suppose function  $\tilde{h} : \mathcal{L} \rightarrow \mathbb{R}$  is totally integral. Then  $\tilde{h}$  is linear on simplex  $\mathcal{L}_\omega$  for each signed ordering  $\omega = (\pi, \sigma)$ .

**Proof.** We use induction on  $n = |V|$ . For  $n = 1$  the claim is straightforward; suppose that  $n \geq 2$ . Consider signed ordering  $\omega = (\pi, \sigma)$ . We need to prove that  $\tilde{h}$  is linear on the boundary  $\partial \mathcal{L}_\omega$ ; this will imply that  $\tilde{h}$  is linear on  $\mathcal{L}_\omega$  since otherwise  $\tilde{h}$  would have an extreme point in the interior  $\mathcal{L}_\omega \setminus \partial \mathcal{L}_\omega$  which cannot be integral.

Let  $X = \{\mathbf{x}^0, \dots, \mathbf{x}^n\}$  be the set of extreme points of  $\mathcal{L}_\omega$  defined by (3). The boundary  $\partial \mathcal{L}_\omega$  is the union of  $n + 1$  facets  $\mathcal{L}_\omega^0, \dots, \mathcal{L}_\omega^n$  where  $\mathcal{L}_\omega^i$  is the convex hull of points in  $X \setminus \{\mathbf{x}^i\}$ . Let us prove that  $\tilde{h}$  is linear on  $\mathcal{L}_\omega^0$ . All points  $\mathbf{x} \in X \setminus \{\mathbf{x}^0\}$  satisfy  $x_1 = \sigma_1$ , therefore  $\mathcal{L}_\omega^0 = \{\mathbf{x} \in \mathcal{L}_\omega \mid x_1 = \sigma_1\}$ . Consider a function of  $n - 1$  variables  $\tilde{h}'(x_2, \dots, x_n) = \tilde{h}(\sigma_1, x_2, \dots, x_n)$ , and let  $\mathcal{L}_\omega^0$  be the projection of  $\mathcal{L}_\omega^0$  to  $\mathbb{R}^{V \setminus \{1\}}$ . By the induction hypothesis  $\tilde{h}'$  is linear on  $\mathcal{L}_\omega^0$ , and thus  $\tilde{h}$  is linear on  $\mathcal{L}_\omega^0$ .

The fact that  $\tilde{h}$  is linear on other facets can be proved in a similar way. Note that for  $i = 2, \dots, n - 1$  there holds  $\mathcal{L}_\omega^i = \{\mathbf{x} \in \mathcal{L}_\omega \mid x_i = \sigma_{i-1} \sigma_i x_{i-1}\}$ , and for  $i = n$  we have  $\mathcal{L}_\omega^n = \{\mathbf{x} \in \mathcal{L}_\omega \mid x_n = 0\}$ .  $\square$

**Corollary 7.** Suppose function  $\tilde{h} : \mathcal{L} \rightarrow \mathbb{R}$  with  $\tilde{h}(\mathbf{0}) = 0$  is totally integral. Let  $h$  be the restriction of  $\tilde{h}$  to  $\mathcal{L}^{1/2}$  and  $\hat{h}$  be the Lovász extension of  $h$ . Then  $\tilde{h}$  and  $\hat{h}$  coincide on  $\mathcal{L}$ .

Theorem 5 and Corollary 7 imply the first part of Theorem 3. The second part will follow from

**Lemma 8.** If  $h : \mathcal{L}^{1/2} \rightarrow \mathbb{R}$  with  $h(\mathbf{0}) = 0$  is bisubmodular then its Lovász extension  $\hat{h} : \mathcal{L} \rightarrow \mathbb{R}$  is totally integral.

**Proof.** We use induction on  $n = |V|$ . For  $n = 1$  the claim is straightforward; suppose that  $n \geq 2$ . By Theorem 5,  $\hat{h}$  is convex on  $\mathcal{L}$ . Function  $\hat{h}$  is integral since it is linear on each simplex  $\mathcal{L}_\omega$  and vertices of  $\mathcal{L}_\omega$  belong to  $\mathcal{L}^{1/2}$ . It remains to show that functions  $\hat{h}'$  considered in Definition 4 are totally integral. Consider the following functions  $h' : \{-1, 0, 1\}^{V \setminus \{n\}} \rightarrow \mathbb{R}$ :

$$\begin{aligned} h'(x_1, \dots, x_{n-1}) &= h(x_1, \dots, x_{n-1}, x_{n-1}) \\ h'(x_1, \dots, x_{n-1}) &= h(x_1, \dots, x_{n-1}, -x_{n-1}) \\ h'(x_1, \dots, x_{n-1}) &= h(x_1, \dots, x_{n-1}, \gamma), \quad \gamma \in \{-1, 0, 1\}. \end{aligned}$$

It can be checked that these functions are bisubmodular, and their Lovász extensions coincide with respective functions  $\hat{h}'$  used in Definition 4. The claim now follows from the induction hypothesis.  $\square$

### 3. A new characterization of bisubmodularity

In this section we give an alternative definition of bisubmodularity; it will be helpful later for describing a relationship to the roof duality. As is often done for bisubmodular functions, we will encode each half-integral value  $x_i \in \{0, 1, \frac{1}{2}\}$  via two Boolean variables  $(u_i, u_{i'})$  according to the following rules:

$$0 \leftrightarrow (0, 1) \quad 1 \leftrightarrow (1, 0) \quad \frac{1}{2} \leftrightarrow (0, 0).$$

Thus, labelings in  $\mathcal{K}^{1/2}$  will be represented via labelings in the set

$$\mathcal{X}^- = \{\mathbf{u} \in \{0, 1\}^{\bar{V}} \mid (u_i, u_{i'}) \neq (1, 1) \forall i \in V\}$$

where  $\bar{V} = \{i, i' \mid i \in V\}$  is a set with  $2n$  nodes. The node  $i'$  for  $i \in V$  is called the “mate” of  $i$ ; intuitively, variable  $u_{i'}$  corresponds to the complement of  $u_i$ . We define  $(i')' = i$  for  $i \in V$ . Labelings in  $\mathcal{X}^-$  will be denoted either by a single letter, e.g.  $\mathbf{u}$  or  $\mathbf{v}$ , or by a pair of letters, e.g.  $(\mathbf{x}, \mathbf{y})$ . In the latter case we assume that the two components correspond to labelings of  $V$  and  $\bar{V} \setminus V$ , respectively, and the order of variables in both components match. Using this convention, the one-to-one

<sup>2</sup> Note, Qi formulates this result slightly differently:  $\hat{h}$  is assumed to be convex on  $\mathbb{R}^V$  rather than on  $\mathcal{L}$ . However, it is easy to see that convexity of  $\hat{h}$  on  $\mathcal{L}$  implies convexity of  $\hat{h}$  on  $\mathbb{R}^V$ . Indeed, it can be checked that  $\hat{h}$  is positively homogeneous, i.e.  $\hat{h}(\gamma \mathbf{x}) = \gamma \hat{h}(\mathbf{x})$  for any  $\gamma \geq 0, \mathbf{x} \in \mathbb{R}^V$ . Therefore, for any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^V$  and  $\alpha, \beta \geq 0$  with  $\alpha + \beta = 1$  there holds

$$\hat{h}(\alpha \mathbf{x} + \beta \mathbf{y}) = \frac{1}{\gamma} \hat{h}(\alpha \gamma \mathbf{x} + \beta \gamma \mathbf{y}) \leq \frac{\alpha}{\gamma} \hat{h}(\gamma \mathbf{x}) + \frac{\beta}{\gamma} \hat{h}(\gamma \mathbf{y}) = \alpha \hat{h}(\mathbf{x}) + \beta \hat{h}(\mathbf{y})$$

where the inequality in the middle follows from convexity of  $\hat{h}$  on  $\mathcal{L}$ , assuming that  $\gamma$  is a sufficiently small constant.

mapping  $\mathcal{X}^- \rightarrow \mathcal{K}^{1/2}$  can be written as  $(\mathbf{x}, \mathbf{y}) \mapsto \frac{1}{2}(\mathbf{x} + \bar{\mathbf{y}})$ . Accordingly, instead of function  $f : \mathcal{K}^{1/2} \rightarrow \mathbb{R}$  we will work with the function  $g : \mathcal{X}^- \rightarrow \mathbb{R}$  defined by

$$g(\mathbf{x}, \mathbf{y}) = f\left(\frac{\mathbf{x} + \bar{\mathbf{y}}}{2}\right). \quad (5)$$

Note that the set of integer labelings  $\mathcal{B} \subset \mathcal{K}^{1/2}$  corresponds to the set  $\mathcal{X}^\circ = \{\mathbf{u} \in \mathcal{X}^- \mid (u_i, u_{i'}) \neq (0, 0)\}$ , so function  $g : \mathcal{X}^- \rightarrow \mathbb{R}$  can be viewed as a discrete relaxation of function  $g : \mathcal{X}^\circ \rightarrow \mathbb{R}$ .

**Definition 9.** Function  $f : \mathcal{X}^- \rightarrow \mathbb{R}$  is called bisubmodular if

$$f(\mathbf{u} \sqcap \mathbf{v}) + f(\mathbf{u} \sqcup \mathbf{v}) \leq f(\mathbf{u}) + f(\mathbf{v}) \quad \forall \mathbf{u}, \mathbf{v} \in \mathcal{X}^- \quad (6)$$

where  $\mathbf{u} \sqcap \mathbf{v} = \mathbf{u} \wedge \mathbf{v}$ ,  $\mathbf{u} \sqcup \mathbf{v} = \text{REDUCE}(\mathbf{u} \vee \mathbf{v})$  and  $\text{REDUCE}(\mathbf{w})$  is the labeling obtained from  $\mathbf{w}$  by changing labels  $(w_i, w_{i'})$  from  $(1, 1)$  to  $(0, 0)$  for all  $i \in V$ .

To describe a new characterization, we need to introduce some additional notation. We denote  $\mathcal{X} = \{0, 1\}^{\bar{V}}$  to be the set of all Boolean labeling of  $\bar{V}$ . For a labeling  $\mathbf{u} \in \mathcal{X}$ , define labeling  $\mathbf{u}'$  by  $(\mathbf{u}')_i = \bar{u}_{i'}$ . Labels  $(u_i, u_{i'})$  are transformed according to the rules

$$(0, 1) \rightarrow (0, 1) \quad (1, 0) \rightarrow (1, 0) \quad (0, 0) \rightarrow (1, 1) \quad (1, 1) \rightarrow (0, 0). \quad (7)$$

Equivalently, this mapping can be written as  $(\mathbf{x}, \mathbf{y})' = (\bar{\mathbf{y}}, \bar{\mathbf{x}})$ . Note that  $\mathbf{u}'' = \mathbf{u}$ ,  $(\mathbf{u} \wedge \mathbf{v})' = \mathbf{u}' \vee \mathbf{v}'$  and  $(\mathbf{u} \vee \mathbf{v})' = \mathbf{u}' \wedge \mathbf{v}'$  for  $\mathbf{u}, \mathbf{v} \in \mathcal{X}$ . Next, we define sets

$$\begin{aligned} \mathcal{X}^- &= \{\mathbf{u} \in \mathcal{X} \mid \mathbf{u} \leq \mathbf{u}'\} = \{\mathbf{u} \in \mathcal{X} \mid (u_i, u_{i'}) \neq (1, 1) \forall i \in \bar{V}\} \\ \mathcal{X}^+ &= \{\mathbf{u} \in \mathcal{X} \mid \mathbf{u} \geq \mathbf{u}'\} = \{\mathbf{u} \in \mathcal{X} \mid (u_i, u_{i'}) \neq (0, 0) \forall i \in \bar{V}\} \\ \mathcal{X}^\circ &= \{\mathbf{u} \in \mathcal{X} \mid \mathbf{u} = \mathbf{u}'\} = \{\mathbf{u} \in \mathcal{X} \mid (u_i, u_{i'}) \in \{(0, 1), (1, 0)\} \forall i \in \bar{V}\} = \mathcal{X}^- \cap \mathcal{X}^+ \\ \mathcal{X}^* &= \mathcal{X}^- \cup \mathcal{X}^+. \end{aligned}$$

Clearly,  $\mathbf{u} \in \mathcal{X}^-$  if and only if  $\mathbf{u}' \in \mathcal{X}^+$ . Also, any function  $g : \mathcal{X}^- \rightarrow \mathbb{R}$  can be uniquely extended to a function  $g : \mathcal{X}^* \rightarrow \mathbb{R}$  so that the following condition holds:

$$g(\mathbf{u}') = g(\mathbf{u}) \quad \forall \mathbf{u} \in \mathcal{X}^*. \quad (8)$$

**Proposition 10.** Let  $g : \mathcal{X}^* \rightarrow \mathbb{R}$  be a function satisfying (8). The following conditions are equivalent:

- (a)  $g$  is bisubmodular, i.e. it satisfies (6).
- (b)  $g$  satisfies the following inequalities:

$$g(\mathbf{u} \wedge \mathbf{v}) + g(\mathbf{u} \vee \mathbf{v}) \leq g(\mathbf{u}) + g(\mathbf{v}) \quad \text{if } \mathbf{u}, \mathbf{v}, \mathbf{u} \wedge \mathbf{v}, \mathbf{u} \vee \mathbf{v} \in \mathcal{X}^*. \quad (9)$$

- (c)  $g$  satisfies those inequalities in (6) for which  $\mathbf{u} = \mathbf{w} \vee \mathbf{e}^i$ ,  $\mathbf{v} = \mathbf{w} \vee \mathbf{e}^j$  where  $\mathbf{w} = \mathbf{u} \wedge \mathbf{v}$  and  $i, j$  are distinct nodes in  $\bar{V}$  with  $w_i = w_j = 0$ . Here  $\mathbf{e}^k$  for node  $k \in \bar{V}$  denotes the labeling in  $\mathcal{X}$  with  $e_k^k = 1$  and  $e_{k'}^k = 0$  for  $k' \in \bar{V} \setminus \{k\}$ .
- (d)  $g$  satisfies those inequalities in (9) for which  $\mathbf{u} = \mathbf{w} \vee \mathbf{e}^i$ ,  $\mathbf{v} = \mathbf{w} \vee \mathbf{e}^j$  where  $\mathbf{w} = \mathbf{u} \wedge \mathbf{v}$  and  $i, j$  are distinct nodes in  $\bar{V}$  with  $z_i = z_j = 0$ .

A proof is given in [Appendix A](#). It is similar to the standard proof of equivalent definitions of submodularity, with a minor complication: in the induction argument for the direction (d)  $\Rightarrow$  (b) we need to make sure that that we move only through the labelings in  $\mathcal{X}^*$ . Note, an equivalent of characterization (c) was given by Ando et al. [2]; we state it here for completeness.

**Remark 1.** In order to compare characterizations (b, d) to existing characterizations (a, c), we need to analyze the sets of inequalities in (b, d) modulo Eq. (8), i.e. after replacing terms  $g(\mathbf{w})$ ,  $\mathbf{w} \in \mathcal{X}^+$  with  $g(\mathbf{w}')$ . It can be seen that the inequalities in (a) are neither a subset nor superset of those in (b),<sup>3</sup> so (b) is a new characterization. It is also possible to show that from this point of view (c) and (d) are equivalent.

<sup>3</sup> Denote  $\mathbf{u} = \begin{pmatrix} 1010 \\ 0000 \end{pmatrix}$  and  $\mathbf{v} = \begin{pmatrix} 0100 \\ 0010 \end{pmatrix}$  where the top and bottom rows correspond to the labelings of  $V$  and  $\bar{V} \setminus V$  respectively, with  $|V| = 4$ . Plugging pair  $(\mathbf{u}, \mathbf{v})$  into (6) gives the following inequality:

$$g\left(\begin{pmatrix} 0000 \\ 0000 \end{pmatrix}\right) + g\left(\begin{pmatrix} 1100 \\ 0000 \end{pmatrix}\right) \leq g\left(\begin{pmatrix} 1010 \\ 0000 \end{pmatrix}\right) + g\left(\begin{pmatrix} 0100 \\ 0010 \end{pmatrix}\right).$$

This inequality is a part of (a), but it is not present in (b): pairs  $(\mathbf{u}, \mathbf{v})$  and  $(\mathbf{u}', \mathbf{v}')$  do not satisfy the RHS of (9), while pairs  $(\mathbf{u}, \mathbf{v}')$  and  $(\mathbf{u}', \mathbf{v})$  give a different inequality:

$$g\left(\begin{pmatrix} 1000 \\ 0000 \end{pmatrix}\right) + g\left(\begin{pmatrix} 0100 \\ 0000 \end{pmatrix}\right) \leq g\left(\begin{pmatrix} 1010 \\ 0000 \end{pmatrix}\right) + g\left(\begin{pmatrix} 0100 \\ 0010 \end{pmatrix}\right)$$

where we used condition (8). Conversely, the second inequality is a part of (b) but it is not present in (a).

#### 4. Submodular relaxations and roof duality

Consider a submodular function  $g : \mathcal{X} \rightarrow \mathbb{R}$  satisfying the following “symmetry” condition:

$$g(\mathbf{u}') = g(\mathbf{u}) \quad \forall \mathbf{u} \in \mathcal{X}. \quad (10)$$

We call such function  $g$  a *submodular relaxation* of function  $f(\mathbf{x}) = g(\mathbf{x}, \bar{\mathbf{x}})$ . Clearly, it satisfies conditions of Proposition 10, so  $g$  is also a bisubmodular relaxation of  $f$ . Furthermore, minimizing  $g$  is equivalent to minimizing its restriction  $g : \mathcal{X}^- \rightarrow \mathbb{R}$ ; indeed, if  $\mathbf{u} \in \mathcal{X}$  is a minimizer of  $g$  then so are  $\mathbf{u}'$  and  $\mathbf{u} \wedge \mathbf{u}' \in \mathcal{X}^-$ .

In this section we will do the following: (i) prove that any pseudo-Boolean function  $f : \mathcal{B} \rightarrow \mathbb{R}$  has a submodular relaxation  $g : \mathcal{X} \rightarrow \mathbb{R}$ ; (ii) show that the roof duality relaxation for quadratic pseudo-Boolean functions is a submodular relaxation, and it dominates all other bisubmodular relaxations; (iii) show that for non-quadratic pseudo-Boolean functions bisubmodular relaxations can be tighter than submodular ones; (iv) prove that similar to the roof duality relaxation, bisubmodular relaxations possess the *persistence* property.

##### 4.1. Review of roof duality

Consider a quadratic pseudo-Boolean function  $f : \mathcal{B} \rightarrow \mathbb{R}$ :

$$f(\mathbf{x}) = \sum_{i \in V} f_i(x_i) + \sum_{(i,j) \in E} f_{ij}(x_i, x_j) \quad (11)$$

where  $(V, E)$  is an undirected graph without self-loops and parallel edges and  $x_i \in \{0, 1\}$  for  $i \in V$  are Boolean variables. Hammer et al. [13] formulated several linear programming relaxations of this function and showed their equivalence. One of these formulations was called a *roof dual*. An efficient maxflow-based method for solving the roof duality relaxation was given by Hammer et al. [5,4].

We will rely on this algorithmic description of the roof duality approach [4]. The method's idea can be summarized as follows. Each variable  $x_i$  is replaced with two Boolean variables  $u_i$  and  $u_{i'}$  corresponding to  $x_i$  and  $1 - x_i$  respectively. The new set of nodes is  $\bar{V} = \{i, i' \mid i \in V\}$ . Next, function  $f$  is transformed to a function  $g : \mathcal{X} \rightarrow \mathbb{R}$  by replacing each term according to the following rules:

$$f_i(x_i) \mapsto \frac{1}{2}[f_i(u_i) + f_i(\bar{u}_{i'})] \quad (12a)$$

$$f_{ij}(x_i, x_j) \mapsto \frac{1}{2}[f_{ij}(u_i, u_j) + f_{ij}(\bar{u}_{i'}, \bar{u}_{j'})] \quad \text{if } f_{ij}(\cdot, \cdot) \text{ is submodular} \quad (12b)$$

$$f_{ij}(x_i, x_j) \mapsto \frac{1}{2}[f_{ij}(u_i, \bar{u}_{j'}) + f_{ij}(\bar{u}_{i'}, u_j)] \quad \text{if } f_{ij}(\cdot, \cdot) \text{ is not submodular} \quad (12c)$$

$g$  is a submodular quadratic pseudo-Boolean function, so it can be minimized via a maxflow algorithm. If  $\mathbf{u} \in \mathcal{X}$  is a minimizer of  $g$  then the roof duality relaxation has a minimizer  $\hat{\mathbf{x}}$  with  $\hat{x}_i = \frac{1}{2}(u_i + \bar{u}_{i'})$  [4].

It is easy to check that  $g(\mathbf{u}) = g(\mathbf{u}')$  for all  $\mathbf{u} \in \mathcal{X}$ , therefore  $g$  is a submodular relaxation. Also,  $f$  and  $g$  are equivalent when  $u_{i'} = \bar{u}_i$  for all  $i \in V$ , i.e.

$$g(\mathbf{x}, \bar{\mathbf{x}}) = f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{B}. \quad (13)$$

*Invariance to variable flipping.* Suppose that  $g$  is a (bi-)submodular relaxation of function  $f : \mathcal{B} \rightarrow \mathbb{R}$ . Let  $i$  be a fixed node in  $V$ , and consider function  $f'(\mathbf{x})$  obtained from  $f(\mathbf{x})$  by a change of coordinates  $x_i \mapsto \bar{x}_i$  and function  $g'(\mathbf{u})$  obtained from  $g(\mathbf{u})$  by swapping variables  $u_i$  and  $u_{i'}$ . It is easy to check that  $g'$  is a (bi-)submodular relaxation of  $f'$ . Furthermore, if  $f$  is a quadratic pseudo-Boolean function and  $g$  is its submodular relaxation constructed by the roof duality approach, then applying the roof duality approach to  $f'$  yields function  $g'$ . We will sometimes use such a “flipping” operation for reducing the number of considered cases.

*Conversion to roof duality* Let us now consider a non-quadratic pseudo-Boolean function  $f : \mathcal{B} \rightarrow \mathbb{R}$ . Several papers [32,1,16] proposed the following scheme: (1) Convert  $f$  to a quadratic pseudo-Boolean function  $\tilde{f}$  by introducing  $k$  auxiliary Boolean variables so that  $f(\mathbf{x}) = \min_{\alpha \in \{0,1\}^k} \tilde{f}(\mathbf{x}, \alpha)$  for all labelings  $\mathbf{x} \in \mathcal{B}$ . (2) Construct submodular relaxation  $\tilde{g}(\mathbf{x}, \alpha, \mathbf{y}, \beta)$  of  $\tilde{f}$  by applying the roof duality relaxation to  $\tilde{f}$ ; then

$$\tilde{g}(\mathbf{x}, \alpha, \mathbf{y}, \beta) = \tilde{g}(\bar{\mathbf{y}}, \bar{\beta}, \bar{\mathbf{x}}, \bar{\alpha}) \quad \tilde{g}(\mathbf{x}, \alpha, \bar{\mathbf{x}}, \bar{\alpha}) = \tilde{f}(\mathbf{x}, \alpha) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{B}, \alpha, \beta \in \{0, 1\}^k.$$

(3) Obtain function  $g$  by minimizing out auxiliary variables:  $g(\mathbf{x}, \mathbf{y}) = \min_{\alpha, \beta \in \{0,1\}^k} \tilde{g}(\mathbf{x}, \alpha, \mathbf{y}, \beta)$ .



One can check that  $g(\mathbf{x}, \mathbf{y}) = g(\bar{\mathbf{y}}, \bar{\mathbf{x}})$ , so  $g$  is a submodular relaxation.<sup>4</sup> In general, however, it may not be a relaxation of function  $f$ , i.e. (13) may not hold; we are only guaranteed to have  $g(\mathbf{x}, \bar{\mathbf{x}}) \leq f(\mathbf{x})$  for all labelings  $\mathbf{x} \in \mathcal{B}$ .

#### 4.2. Submodular and bisubmodular relaxations

We are now ready to present results (i)–(iv) formulated in the beginning of Section 4.

**Existence of submodular relaxations.** It is easy to check that if  $f : \mathcal{B} \rightarrow \mathbb{R}$  is submodular then function  $g(\mathbf{x}, \mathbf{y}) = \frac{1}{2}[f(\mathbf{x}) + f(\bar{\mathbf{y}})]$  is a submodular relaxation of  $f$ .<sup>5</sup> Thus, monomials of the form  $c\Pi_{i \in A} x_i$  where  $c \leq 0$  and  $A \subseteq V$  have submodular relaxations. Using the “flipping” operation  $x_i \mapsto \bar{x}_i$ , we conclude that submodular relaxations also exist for monomials of the form  $c\Pi_{i \in A} x_i \Pi_{i \in B} \bar{x}_i$  where  $c \leq 0$  and  $A, B$  are disjoint subsets of  $U$ . It is known that any pseudo-Boolean function  $f$  can be represented as a sum of such monomials (see e.g. [4]; we need to represent  $-f$  as a *posiform* and take its negative). This implies that any pseudo-Boolean function  $f$  has a submodular relaxation.

Note that this argument is due to Lu and Williams [24] who converted function  $f$  to a sum of monomials of the form  $c\Pi_{i \in A} x_i$  and  $c\bar{x}_k \Pi_{i \in A} x_i$ ,  $c \leq 0$ ,  $k \notin A$ . It is possible to show that the relaxation proposed in [24] is equivalent to the submodular relaxation constructed by the scheme above (we omit the derivation).

**Submodular vs. bisubmodular relaxations.** An important question is whether bisubmodular relaxations are more “powerful” compared to submodular ones. The next theorem gives a class of functions for which the answer is negative; its proof is given in Appendix B.

**Theorem 11.** *Let  $g$  be the submodular relaxation of a quadratic pseudo-Boolean function  $f$  defined by (12). Then  $g$  dominates any other bisubmodular relaxation  $\bar{g}$  of  $f$ , i.e.  $g(\mathbf{u}) \geq \bar{g}(\mathbf{u})$  for all  $\mathbf{u} \in \mathcal{X}^-$ .*

For non-quadratic pseudo-Boolean functions, however, the situation can be different. In Appendix C we give an example of a function  $f$  of  $n = 4$  variables which has a tight bisubmodular relaxation  $g$  (i.e.  $g$  has a minimizer in  $\mathcal{X}^\circ$ ), but all submodular relaxations are not tight.

**Persistency.** Finally, we show that bisubmodular functions possess the *autarky* property, which implies *persistency*.

**Proposition 12.** *Let  $f : \mathcal{X}^{1/2} \rightarrow \mathbb{R}$  be a bisubmodular function and  $\mathbf{x} \in \mathcal{X}^{1/2}$  be its minimizer.*

[Autarky] *Let  $\mathbf{y}$  be a labeling in  $\mathcal{B}$ . Consider labeling  $\mathbf{z} = (\mathbf{y} \sqcup \mathbf{x}) \sqcup \mathbf{x}$ . Then  $\mathbf{z} \in \mathcal{B}$  and  $f(\mathbf{z}) \leq f(\mathbf{y})$ .*

[Persistency] *Function  $f : \mathcal{B} \rightarrow \mathbb{R}$  has a minimizer  $\mathbf{x}^* \in \mathcal{B}$  such that  $x_i^* = x_i$  for nodes  $i \in V$  with integral  $x_i$ .*

**Proof.** It can be checked that  $z_i = y_i$  if  $x_i = \frac{1}{2}$  and  $z_i = x_i$  if  $x_i \in \{0, 1\}$ . Thus,  $\mathbf{z} \in \mathcal{B}$ . For any  $\mathbf{w} \in \mathcal{X}^{1/2}$  there holds  $f(\mathbf{w} \sqcup \mathbf{x}) \leq f(\mathbf{w}) + [f(\mathbf{x}) - f(\mathbf{w} \sqcap \mathbf{x})] \leq f(\mathbf{w})$ . This implies that  $f((\mathbf{y} \sqcup \mathbf{x}) \sqcup \mathbf{x}) \leq f(\mathbf{y})$ . Applying the autarky property to a labeling  $\mathbf{y} \in \arg \min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{B}\}$  yields persistency.  $\square$

## 5. Conclusions and future work

We showed that bisubmodular functions provide a natural generalization of the roof duality approach to higher-order terms. This can be viewed as a non-submodular analogue of the fact that submodular functions generalize the  $s$ – $t$  minimum cut problem with non-negative weights to higher-order terms.

As mentioned in the introduction, this work has been motivated by computer vision applications that use functions of the form  $f(\mathbf{x}) = \sum_C f_C(\mathbf{x})$ . An important open question is how to construct bisubmodular relaxations  $\hat{f}_C$  for individual terms. For terms of low order, e.g. with  $|C| = 3$ , this potentially could be done by solving a small linear program.

Another important question is how to minimize such functions. Algorithms in [12,25] are unlikely to be practical for most vision problems, which typically have tens of thousands of variables. However, in our case we need to minimize a bisubmodular function which has a special structure: it is represented as a sum of low-order bisubmodular terms. We recently showed [20] that a sum of low-order *submodular* terms can be optimized more efficiently using maxflow-like techniques. We conjecture that similar techniques can be developed for bisubmodular functions as well.

<sup>4</sup> It is well-known that minimizing variables out preserves submodularity. Indeed, suppose that  $h(\mathbf{x}) = \min_{\alpha} \tilde{h}(\mathbf{x}, \alpha)$  where  $\tilde{h}$  is a submodular function. Then  $h$  is also submodular since

$$h(\mathbf{x}) + h(\mathbf{y}) = \tilde{h}(\mathbf{x}, \alpha) + \tilde{h}(\mathbf{y}, \beta) \geq \tilde{h}(\mathbf{x} \wedge \mathbf{y}, \alpha \wedge \beta) + \tilde{h}(\mathbf{x} \vee \mathbf{y}, \alpha \vee \beta) \geq h(\mathbf{x} \wedge \mathbf{y}) + h(\mathbf{x} \vee \mathbf{y}).$$

<sup>5</sup> In fact, it dominates all other bisubmodular relaxations  $\bar{g} : \mathcal{X}^- \rightarrow \mathbb{R}$  of  $f$ . Indeed, consider labeling  $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^-$ . It can be checked that  $(\mathbf{x}, \mathbf{y}) = \mathbf{u} \sqcap \mathbf{v} = \mathbf{u} \sqcup \mathbf{v}$  where  $\mathbf{u} = (\mathbf{x}, \bar{\mathbf{x}})$  and  $\mathbf{v} = (\bar{\mathbf{y}}, \mathbf{y})$ , therefore  $\bar{g}(\mathbf{x}, \mathbf{y}) \leq \frac{1}{2}[\bar{g}(\mathbf{u}) + \bar{g}(\mathbf{v})] = \frac{1}{2}[f(\mathbf{x}) + f(\bar{\mathbf{y}})] = g(\mathbf{x}, \mathbf{y})$ .

## Appendix A. Proof of Proposition 10 (definitions of bisubmodularity)

Directions (a)  $\Rightarrow$  (c) and (b)  $\Rightarrow$  (d) are trivial. Below we prove directions (b)  $\Rightarrow$  (a), (d)  $\Rightarrow$  (b) and (c)  $\Rightarrow$  (d). We use the following notation: for a labeling  $\mathbf{u} \in \mathcal{X}$  and distinct nodes  $i, j \in \bar{V}$  we denote  $[\mathbf{u}]_i = (u_i, u_{i'})$ ,  $[\mathbf{u}]_{ij} = (u_i, u_{i'}, u_j, u_{j'})$ .

*Direction (b)  $\Rightarrow$  (a).* For labelings  $\mathbf{u}, \mathbf{v} \in \mathcal{X}^-$  define  $\alpha = \mathbf{u} \wedge \mathbf{v}'$ ,  $\beta = \mathbf{u}' \wedge \mathbf{v}$ . Clearly,  $\alpha, \beta \in \mathcal{X}^-$ . Also,  $\alpha \wedge \beta = \mathbf{u} \sqcap \mathbf{v}$  and  $\alpha \vee \beta = \mathbf{u} \sqcup \mathbf{v}$ . We can write

$$\begin{aligned} g(\mathbf{u} \sqcap \mathbf{v}) + g(\mathbf{u} \sqcup \mathbf{v}) &= g(\alpha \wedge \beta) + g(\alpha \vee \beta) \leq g(\alpha) + g(\beta) = g(\alpha) + g(\beta') \\ &= g(\mathbf{u} \wedge \mathbf{v}') + g(\mathbf{u} \vee \mathbf{v}') \leq g(\mathbf{u}) + g(\mathbf{v}') = g(\mathbf{u}) + g(\mathbf{v}) \end{aligned}$$

where we used conditions (8) and (9). (It can be checked that all labelings involved belong to  $\mathcal{X}^*$ .)

*Direction (d)  $\Rightarrow$  (b).* We show that all inequalities in (9) hold using induction on the Hamming distance  $\|\mathbf{u} - \mathbf{v}\|_1 = \sum_{i \in V} |u_i - v_i|$  between  $\mathbf{u}$  and  $\mathbf{v}$ . The base case  $\|\mathbf{u} - \mathbf{v}\|_1 \leq 2$  is straightforward: if labelings  $\mathbf{u}, \mathbf{v}, \mathbf{u} \vee \mathbf{v}, \mathbf{u} \wedge \mathbf{v} \in \mathcal{X}^*$  are all distinct then (9) follows directly from condition (d), otherwise (9) is an equality.

Suppose  $\|\mathbf{u} - \mathbf{v}\|_1 \geq 3$  and  $\mathbf{u}, \mathbf{v}, \mathbf{u} \vee \mathbf{v}, \mathbf{u} \wedge \mathbf{v} \in \mathcal{X}^*$ . We can assume by symmetry that  $u_i = 1, v_i = 0$  for at least two nodes  $i \in V$ . Among such nodes, let us choose  $i$  as follows: (a) if there is such  $i$  with  $u_{i'} = 1$  pick this node; (b) otherwise if there is such  $i$  with  $v_{i'} = 1$  pick this node; (c) otherwise pick an arbitrary such node. Let  $\tilde{\mathbf{u}}$  be the labeling obtained from  $\mathbf{u}$  by switching the label of  $i$  from 1 to 0. It can be checked that  $\mathbf{u} \vee (\tilde{\mathbf{u}} \vee \mathbf{v}) = \mathbf{u} \vee \mathbf{v}$  and  $\mathbf{u} \wedge (\tilde{\mathbf{u}} \vee \mathbf{v}) = \tilde{\mathbf{u}}$ . There holds  $\|\mathbf{u} - (\tilde{\mathbf{u}} \vee \mathbf{v})\|_1 = |\{j \in V \mid u_j < v_j\}| + 1 < |\{j \in V \mid u_j < v_j\}| + |\{j \in V \mid u_j > v_j\}| = \|\mathbf{u} - \mathbf{v}\|_1$  and  $\|\tilde{\mathbf{u}} - \mathbf{v}\|_1 = \|\mathbf{u} - \mathbf{v}\|_1 - 1$ , so by the induction hypothesis

$$g(\mathbf{u} \vee \mathbf{v}) - g(\mathbf{u}) \leq g(\tilde{\mathbf{u}} \vee \mathbf{v}) - g(\tilde{\mathbf{u}}) \leq g(\mathbf{v}) - g(\tilde{\mathbf{u}} \wedge \mathbf{v}) = g(\mathbf{v}) - g(\mathbf{u} \wedge \mathbf{v})$$

provided that all labelings involved belong to  $\mathcal{X}^*$ . This fact is proven below.

Let us show that  $\tilde{\mathbf{u}} \in \mathcal{X}^*$ . If  $u_{i'} = 1$  then  $[\mathbf{u}]_i = (1, 1)$ ,  $\tilde{\mathbf{u}}$  is obtained from  $\mathbf{u}$  by switching  $[\mathbf{u}]_i$  from  $(1, 1)$  to  $(0, 1)$ , so  $\mathbf{u} \in \mathcal{X}^*$  implies  $\tilde{\mathbf{u}} \in \mathcal{X}^*$ . Suppose that  $u_{i'} = 0$ ; this means that  $i$  was not selected in rule (a). If  $\mathbf{u} \notin \mathcal{X}^-$  then there exists  $j \in V$  with  $[\mathbf{u}]_j = (1, 1)$ . Since case (a) was not “triggered”, we must have  $[\mathbf{v}]_j = (1, 1)$ . But then  $[\mathbf{u} \wedge \mathbf{v}]_{ij} = (0, 0, 1, 1)$  so  $\mathbf{u} \wedge \mathbf{v} \notin \mathcal{X}^*$ —a contradiction. Thus,  $\mathbf{u} \in \mathcal{X}^-$  and so  $\tilde{\mathbf{u}} \in \mathcal{X}^-$ .

Let us now show that  $\tilde{\mathbf{w}} = \tilde{\mathbf{u}} \vee \mathbf{v} \in \mathcal{X}^*$ . Clearly,  $\tilde{\mathbf{w}}$  is obtained from  $\mathbf{w} = \mathbf{u} \vee \mathbf{v}$  by switching label  $w_i$  from 1 to 0. If  $\tilde{w}_{i'} = 1$  then  $[\mathbf{w}]_i = (1, 1)$  switches to  $[\tilde{\mathbf{w}}]_i = (0, 1)$ , so  $\mathbf{w} \in \mathcal{X}^*$  implies  $\tilde{\mathbf{w}} \in \mathcal{X}^*$ . Suppose that  $\tilde{w}_{i'} = 0$ , then  $u_{i'} = v_{i'} = 0$ ; this means that rules (a) and (b) were not “triggered”. Let us prove that  $\mathbf{w} \in \mathcal{X}^-$ ; this will imply  $\tilde{\mathbf{w}} \in \mathcal{X}^-$ . Suppose not, then there exist  $j \in V$  with  $[\mathbf{w}]_j = (1, 1)$ . The case  $[\mathbf{v}]_j = (1, 1)$  is impossible since then we would have  $[\mathbf{v}]_{ij} = (0, 0, 1, 1)$  and  $\mathbf{v} \notin \mathcal{X}^*$ —a contradiction. Thus, we can assume by symmetry that  $v_{i'} = 0$ , so  $u_{i'} = 1$ . We must have  $u_j = 1$  or  $v_j = 1$ , which means that either rule (a) or (b) would be triggered—a contradiction.

*Direction (c)  $\Rightarrow$  (d).* Let  $\mathbf{u}, \mathbf{v}$  be labelings with the properties of condition (d). Thus,  $\mathbf{u} = \mathbf{w} \vee \mathbf{e}^i$ ,  $\mathbf{v} = \mathbf{w} \vee \mathbf{e}^j$  and  $w_i = w_j = 0$ .

Suppose that  $j = i'$ , so  $[\mathbf{w}]_i = (0, 0)$ . We must have  $[\mathbf{w}]_k \in \{(0, 1), (1, 0)\}$  for all  $k \in V \setminus \{i, j\}$ , otherwise we would have either  $\mathbf{u} \wedge \mathbf{v} \notin \mathcal{X}^*$  or  $\mathbf{u} \vee \mathbf{v} \notin \mathcal{X}^*$ . Therefore,  $\mathbf{u}, \mathbf{v} \in \mathcal{X}^\circ$  and  $g(\mathbf{u} \sqcup \mathbf{v}) = g(\mathbf{w}) = g(\mathbf{w}') = g(\mathbf{u} \vee \mathbf{v})$ , so (9) follows from (6).

We now assume that  $j \neq i$  and  $j \neq i'$ . We have  $[\mathbf{w}]_{ij} = (0, ?, 0, ?)$ . Cases  $[\mathbf{w}]_{ij} = (0, 0, 0, 1)$  and  $[\mathbf{w}]_{ij} = (0, 1, 0, 0)$  are impossible since then either  $\mathbf{u}$  or  $\mathbf{v}$  would not belong to  $\mathcal{X}^*$ . If  $[\mathbf{w}]_{ij} = (0, 0, 0, 0)$  then  $\mathbf{u}, \mathbf{v} \in \mathcal{X}^-$  and  $\mathbf{u} \sqcup \mathbf{v} = \mathbf{u} \vee \mathbf{v}$ , so (9) follows from (6). It remains to consider the case  $[\mathbf{w}]_{ij} = [\mathbf{w}']_{ij} = (0, 1, 0, 1)$ . Labeling  $\mathbf{u}'$  is obtained from  $\mathbf{w}'$  by switching the label of node  $i'$  from 1 to 0. Similarly,  $\mathbf{v}'$  is obtained from  $\mathbf{w}'$  by switching the label of node  $j'$  from 1 to 0. We have  $[\mathbf{u}']_{ij} = (0, 0, 0, 1)$ ,  $[\mathbf{v}']_{ij} = (0, 1, 0, 0)$ , so  $\mathbf{u}', \mathbf{v}' \in \mathcal{X}^-$ . Furthermore,  $\mathbf{u}' \sqcup \mathbf{v}' = \mathbf{u}' \vee \mathbf{v}'$ . Therefore,

$$g(\mathbf{u} \vee \mathbf{v}) + g(\mathbf{u} \wedge \mathbf{v}) = g(\mathbf{u}' \wedge \mathbf{v}') + g(\mathbf{u}' \vee \mathbf{v}') = g(\mathbf{u}' \sqcap \mathbf{v}') + g(\mathbf{u}' \sqcup \mathbf{v}') \leq g(\mathbf{u}') + g(\mathbf{v}') = g(\mathbf{u}) + g(\mathbf{v}).$$

## Appendix B. Proof of Theorem 11

In order to simplify the proof (i.e. reduce the number of considered cases) we will use the “flipping” operation described in Section 4.

For  $\mathbf{u} \in \mathcal{X}$  we define sets  $V_{00}[\mathbf{u}] = \{i \in V \mid (u_i, u_{i'}) = (0, 0)\}$ ,  $\bar{V}_{00}[\mathbf{u}] = \{i \in \bar{V} \mid (u_i, u_{i'}) = (0, 0)\}$ . In this appendix we denote  $\mathbf{u}^i$  to be the labeling obtained from labeling  $\mathbf{u} \in \mathcal{X}$  by setting the label of node  $i \in \bar{V}$  to 1, i.e.  $\mathbf{u}^i = \mathbf{u} \vee \mathbf{e}^i$ . Similarly, we denote  $\mathbf{u}^{ij} = \mathbf{u} \vee \mathbf{e}^i \vee \mathbf{e}^j$ .

**Lemma 13.** Suppose that  $\mathbf{u} \in \mathcal{X}$  and  $i, j$  are distinct nodes in  $\bar{V}_{00}[\mathbf{u}]$  satisfying the following conditions: (i) if  $i, j \in V$  then the term  $f_{ij}(\cdot, \cdot)$  (if it exists) is non-submodular; (ii) if  $i, j \in \bar{V} \setminus V$  then the term  $f_{i'j'}(\cdot, \cdot)$  (if it exists) is non-submodular; (iii) if  $i \in V, j \in \bar{V} \setminus V$  then the term  $f_{ij'}(\cdot, \cdot)$  (if it exists) is submodular; (iv) if  $i \in \bar{V} \setminus V, j \in V$  then the term  $f_{i'j}(\cdot, \cdot)$  (if it exists) is submodular. Then

$$g(\mathbf{u}) + g(\mathbf{u}^{ij}) = g(\mathbf{u}^i) + g(\mathbf{u}^j). \quad (14)$$



**Proof.** It suffices to prove the lemma in the case when function  $f$  in Eq. (11) has a single term; the general case will then follow by linearity. If this term does not involve nodes  $i/i'$  and  $j/j'$  then the claim is trivial since then  $g(\mathbf{u})$  does not depend on  $u_i$  or  $u_j$ . Thus, we consider terms involving at least one of the nodes  $i, i', j, j'$ .

Suppose that  $j = i'$ ; without loss of generality we can assume that  $i \in V$ . If  $f(\mathbf{x}) = f_i(x_i)$  then the LHS and the RHS of (14) equals  $f_i(0) + f_i(1)$ . If  $f(\mathbf{x}) = f_{ik}(x_i, x_k)$  and term  $f_{ik}(\cdot, \cdot)$  is submodular then the LHS and the RHS of (14) equal  $\frac{1}{2}[f_{ik}(0, u_k) + f_{ik}(1, \bar{u}_k) + f_{ik}(1, u_k) + f_{ik}(0, \bar{u}_k)]$ . The case when  $f(\mathbf{x}) = f_{ik}(x_i, x_k)$  and term  $f_{ik}(\cdot, \cdot)$  is non-submodular can be reduced to the previous one by flipping node  $k$ .

Now suppose that  $j \neq i'$ . By assumption,  $(u_i, u_{i'}, u_j, u_{j'}) = (0, 0, 0, 0)$ . Using flipping, we can ensure that  $i, j \in V$ . (Note that flipping  $i$  and/or  $j$  preserves conditions (i)–(iv).) Suppose  $f(\mathbf{x})$  involves exactly one of the nodes  $i, j$ , say node  $i$ . If  $f(\mathbf{x}) = f_i(x_i)$  then the LHS and the RHS of (14) equal  $\frac{1}{2}[f_i(0) + 3f_i(1)]$ . If  $f(\mathbf{x}) = f_{ik}(x_i, x_k)$  and term  $f_{ik}(\cdot, \cdot)$  is submodular then the LHS and the RHS of (14) equal  $\frac{1}{2}[f_{ik}(0, u_k) + f_{ik}(1, \bar{u}_k) + f_{ik}(1, u_k) + f_{ik}(0, \bar{u}_k)]$ . The case when  $f(\mathbf{x}) = f_{ik}(x_i, x_k)$  and term  $f_{ik}(\cdot, \cdot)$  is non-submodular can be reduced to the previous one by flipping node  $k$ . It remains to consider the case when  $f(\mathbf{x}) = f_{ij}(x_i, x_j)$ . By the lemma's assumption, term  $f_{ij}(\cdot, \cdot)$  is non-submodular, so the LHS and the RHS of (14) equal  $\frac{1}{2}[f_{ij}(0, 1) + f_{ij}(1, 0) + 2f_{ij}(1, 1)]$ .  $\square$

**Lemma 14.** Labeling  $\mathbf{u} \in \mathcal{X}^-$  with  $V_{00}[\mathbf{u}] = \{i\}$  satisfies  $\bar{g}(\mathbf{u}) \leq g(\mathbf{u})$ .

**Proof.** We have  $2\bar{g}(\mathbf{u}) \stackrel{(1)}{\leq} \bar{g}(\mathbf{u}^i) + \bar{g}(\mathbf{u}^{i'}) \stackrel{(2)}{=} g(\mathbf{u}^i) + g(\mathbf{u}^{i'}) \stackrel{(3)}{=} g(\mathbf{u}) + g(\mathbf{u}^{i''}) \stackrel{(4)}{=} 2g(\mathbf{u})$  where (1) holds since  $\bar{g}$  is bisubmodular and  $\mathbf{u}^i \sqcap \mathbf{u}^{i'} = \mathbf{u}^i \sqcup \mathbf{u}^{i'} = \mathbf{u}$ , (2) holds since  $g$  and  $\bar{g}$  are relaxations of  $f$  and  $\mathbf{u}^i, \mathbf{u}^{i'} \in \mathcal{X}^\circ$ , (3) holds by Lemma 13, and (4) holds since  $g(\mathbf{u}^{i''}) = g(\mathbf{u}^i) = g(\mathbf{u})$ .  $\square$

**Lemma 15.** There holds  $\bar{g}(\mathbf{u}) - \bar{g}(\mathbf{u}^i) \leq g(\mathbf{u}) - g(\mathbf{u}^i)$  for all  $\mathbf{u} \in \mathcal{X}^-$  and  $i \in \bar{V}_{00}[\mathbf{u}]$ .

**Proof.** We use induction on  $|\bar{V}_{00}[\mathbf{u}]|$ . If  $\bar{V}_{00}[\mathbf{u}] = \emptyset$  then the claim is trivial. If  $\bar{V}_{00}[\mathbf{u}] = \{i, i'\}$  then the claim follows from Lemma 14 and the fact that  $\bar{g}(\mathbf{u}^i) = g(\mathbf{u}^i)$  which holds since  $\mathbf{u}^i \in \mathcal{X}^\circ$ . Now suppose that there exists  $j \in \bar{V}_{00}[\mathbf{u}] \setminus \{i, i'\}$ . We can assume without loss of generality that labeling  $\mathbf{u}$  and nodes  $i, j$  satisfy conditions of Lemma 13. (If not, we can replace  $j$  with  $j'$ .) We can write

$$\bar{g}(\mathbf{u}) - \bar{g}(\mathbf{u}^i) \stackrel{(1)}{\leq} \bar{g}(\mathbf{u}^j) - \bar{g}(\mathbf{u}^{ij}) \stackrel{(2)}{\leq} g(\mathbf{u}^j) - g(\mathbf{u}^{ij}) \stackrel{(3)}{=} g(\mathbf{u}) - g(\mathbf{u}^i)$$

where (1) holds since  $\bar{g}$  is bisubmodular, (2) holds by the induction hypothesis and (3) follows from Lemma 13.  $\square$

We are now ready to prove Theorem 11, i.e. that  $\bar{g}(\mathbf{u}) \leq g(\mathbf{u})$  for any  $\mathbf{u} \in \mathcal{X}^-$ . We use induction on  $|\bar{V}_{00}[\mathbf{u}]|$ . The case  $V_{00}[\mathbf{u}] = \emptyset$  is trivial and the case  $V_{00}[\mathbf{u}] = \{i\}$  follows from Lemma 14. Suppose that there exists  $j \in \bar{V}_{00}[\mathbf{u}] \setminus \{i, i'\}$ . As before, we can assume without loss of generality that labeling  $\mathbf{u}$  and nodes  $i, j$  satisfy conditions of Lemma 13. We can write

$$\bar{g}(\mathbf{u}) \stackrel{(1)}{\leq} \bar{g}(\mathbf{u}^i) + [\bar{g}(\mathbf{u}^i) - \bar{g}(\mathbf{u}^{ij})] \stackrel{(2)}{\leq} g(\mathbf{u}^i) + [g(\mathbf{u}^i) - g(\mathbf{u}^{ij})] \stackrel{(3)}{=} g(\mathbf{u})$$

where (1) holds since  $\bar{g}$  is bisubmodular, (2) holds by the induction hypothesis and Lemma 14, and (3) follows from Lemma 13.

## Appendix C. Examples of bisubmodular functions

First, let us consider *cardinality-dependent* functions  $g : \mathcal{X}^- \rightarrow \mathbb{R}$ , i.e. functions which can be expressed as

$$g(\mathbf{u}) = G(n_{01}[\mathbf{u}], n_{10}[\mathbf{u}])$$

where  $n_{\alpha\beta}[\mathbf{u}] = |\{i \in V \mid (u_i, u_{i'}) = (\alpha, \beta)\}|$  for a labeling  $\mathbf{u} \in \mathcal{X}$  and function  $G$  is defined over  $D_n = \{(a, b) \in \mathbb{Z}^2 \mid a, b \geq 0, a + b \leq n\}$ . Using Proposition 10(c) it is easy to check that  $f$  is bisubmodular if and only if  $G$  satisfies the following:

$$G(a, b) + G(a - 2, b) \leq 2G(a - 1, b) \quad \forall (a, b) \in D_n, a \geq 2 \quad (15a)$$

$$G(a, b) + G(a, b - 2) \leq 2G(a, b - 1) \quad \forall (a, b) \in D_n, b \geq 2 \quad (15b)$$

$$G(a, b) + G(a - 1, b - 1) \leq G(a - 1, b) + G(a, b - 1) \quad \forall (a, b) \in D_n, a \geq 1, b \geq 1 \quad (15c)$$

$$2G(a, b) \leq G(a + 1, b + 1) \quad \forall (a, b) \in D_n, a + b = n - 1. \quad (15d)$$

**Proposition 16.** Consider function  $G : D_3 \rightarrow \mathbb{R}$  defined by Fig. 1(b). Let  $g$  be the corresponding cardinality-dependent function of  $2n = 6$  Boolean variables. Then (a)  $g$  is bisubmodular and (b) it cannot be extended to a submodular relaxation  $\mathcal{X} \rightarrow \mathbb{R}$ .

**Proof.** Verifying that function  $G$  satisfies (15) is straightforward, so we focus on the claim (b). Suppose a submodular relaxation  $\bar{g} : \mathcal{X} \rightarrow \mathbb{R}$  that extends  $g$  does exist. Without loss of generality we can assume that  $\bar{g}(\mathbf{u}) = \bar{G}(n_{01}[\mathbf{u}], n_{10}[\mathbf{u}], n_{00}[\mathbf{u}], n_{11}[\mathbf{u}])$  for some function  $\bar{G}$  over  $\bar{D}_3 = \{(a, b, c, d) \in \mathbb{Z}^4 \mid a, b, c, d \geq 0, a + b + c + d = 3\}$ . Indeed, let  $\Pi$  be the set of  $3! = 6$  permutations of  $V$ . Any permutation  $\pi \in \Pi$  defines a mapping  $\psi_\pi : \mathcal{X} \rightarrow \mathcal{X}$  in a natural way. Define function  $\bar{g}_\pi : \mathcal{X} \rightarrow \mathbb{R}$  by  $\bar{g}_\pi(\mathbf{u}) = \bar{g}(\psi_\pi(\mathbf{u}))$ . Clearly,  $\bar{g}_\pi$  is also a submodular relaxation extending  $g$ ,

<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>																																
$F(0, 3)$ $F(0, 2) \quad F(1, 2)$ $F(0, 1) \quad F(1, 1) \quad F(2, 1)$ $F(0, 0) \quad F(1, 0) \quad F(2, 0) \quad F(3, 0)$	$0$ $0 \quad 0$ $0 \quad 0 \quad 0$ $-1 \quad 0 \quad 1 \quad 2$	$0$ $-\frac{1}{3} \quad 0$ $-\frac{2}{3} \quad 0 \quad 0$ $-1 \quad 0 \quad 1 \quad 2$	<table border="1"> <tr><td>0000</td><td>3</td><td>1000</td><td>1</td></tr> <tr><td>0001</td><td>2</td><td>1001</td><td>3</td></tr> <tr><td>0010</td><td>4</td><td>1010</td><td>0</td></tr> <tr><td>0011</td><td>10</td><td>1011</td><td>12</td></tr> <tr><td>0100</td><td>2</td><td>1100</td><td>7</td></tr> <tr><td>0101</td><td>12</td><td>1101</td><td>10</td></tr> <tr><td>0110</td><td>13</td><td>1110</td><td>12</td></tr> <tr><td>0111</td><td>12</td><td>1111</td><td>14</td></tr> </table>	0000	3	1000	1	0001	2	1001	3	0010	4	1010	0	0011	10	1011	12	0100	2	1100	7	0101	12	1101	10	0110	13	1110	12	0111	12	1111	14
0000	3	1000	1																																
0001	2	1001	3																																
0010	4	1010	0																																
0011	10	1011	12																																
0100	2	1100	7																																
0101	12	1101	10																																
0110	13	1110	12																																
0111	12	1111	14																																

**Fig. 1.** Examples of bisubmodular functions. (a–c) Cardinality-dependent functions  $g : \mathcal{X}^- \rightarrow \mathbb{R}$  written as  $g(\mathbf{u}) = G(n_{01}[\mathbf{u}], n_{10}[\mathbf{u}])$  where  $n_{\alpha\beta}[\mathbf{u}] = |\{i \in V \mid (u_i, u_{i'}) = (\alpha, \beta)\}|$ . Here  $n = 3$ . (a) Convention for displaying function  $G$ . (b, c) Bisubmodular relaxations of the same function  $f$ . Function (c) can be extended to a submodular relaxation, while (b) cannot be extended. (d) Function  $f$  of  $n = 4$  variables which has a tight bisubmodular relaxation, but all submodular relaxations are not tight.

and therefore so is the function  $\bar{g}_* : \mathcal{X} \rightarrow \mathbb{R}$  given by  $\bar{g}_*(\mathbf{u}) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \bar{g}_\pi(\mathbf{u})$ . Clearly,  $\bar{g}_*(\mathbf{u})$  depends only on the counts  $n_{01}[\mathbf{u}], n_{10}[\mathbf{u}], n_{00}[\mathbf{u}], n_{11}[\mathbf{u}]$ .

If  $c = 0$  or  $d = 0$  for  $(a, b, c, d) \in \bar{D}_3$  then  $\bar{G}(a, b, c, d) = G(a, b)$ . Thus, there are 4 unknown values:  $\bar{G}(0, 1, 1, 1)$ ,  $\bar{G}(1, 0, 1, 1)$ ,  $\bar{G}(0, 0, 1, 2)$ ,  $\bar{G}(0, 0, 2, 1)$ . We will write labelings in  $\mathcal{X}$  as  $\binom{u_i \ u_j \ u_k}{u_{i'} \ u_{j'} \ u_{k'}}$  where  $\{i, j, k\} = U$ . From submodularity of  $\bar{g}$  we get

$$\begin{aligned} \bar{g}\left(\binom{101}{001}\right) + \bar{g}\left(\binom{111}{011}\right) &\leq \bar{g}\left(\binom{101}{011}\right) + \bar{g}\left(\binom{111}{001}\right) \Rightarrow \bar{G}(0, 1, 1, 1) + 0 \leq 0 + 0 \\ \bar{g}\left(\binom{101}{000}\right) + \bar{g}\left(\binom{101}{011}\right) &\leq \bar{g}\left(\binom{101}{001}\right) + \bar{g}\left(\binom{101}{010}\right) \Rightarrow 0 + 0 \leq \bar{G}(0, 1, 1, 1) + 0 \\ \bar{g}\left(\binom{001}{101}\right) + \bar{g}\left(\binom{011}{111}\right) &\leq \bar{g}\left(\binom{001}{111}\right) + \bar{g}\left(\binom{011}{101}\right) \Rightarrow \bar{G}(1, 0, 1, 1) + 0 \leq 1 + 0 \\ \bar{g}\left(\binom{000}{101}\right) + \bar{g}\left(\binom{011}{101}\right) &\leq \bar{g}\left(\binom{001}{101}\right) + \bar{g}\left(\binom{010}{101}\right) \Rightarrow 1 + 0 \leq \bar{G}(1, 0, 1, 1) + 0 \end{aligned}$$

which implies  $\bar{G}(0, 1, 1, 1) = 0$ ,  $\bar{G}(1, 0, 1, 1) = 1$ . Additional submodularity inequalities lead to an inconsistency:

$$\begin{aligned} \bar{g}\left(\binom{001}{011}\right) + \bar{g}\left(\binom{111}{011}\right) &\leq \bar{g}\left(\binom{011}{011}\right) + \bar{g}\left(\binom{101}{011}\right) \Rightarrow 1 + 0 \leq \bar{G}(0, 0, 1, 2) + 0 \\ \bar{g}\left(\binom{011}{000}\right) + \bar{g}\left(\binom{011}{011}\right) &\leq \bar{g}\left(\binom{011}{001}\right) + \bar{g}\left(\binom{011}{010}\right) \Rightarrow 0 + \bar{G}(0, 0, 1, 2) \leq 0 + 0. \quad \square \end{aligned}$$

It should be said that in this particular example  $f$  has a submodular relaxation  $\tilde{g}$  with the same minimum as  $g$  (the restriction of  $\tilde{g}$  to  $\mathcal{X}^-$  is shown in Fig. 1(c)). Although  $\tilde{g}(\mathbf{u}) < g(\mathbf{u})$  for some  $\mathbf{u} \in \mathcal{X}^-$ , both functions attain the minimum of  $-1$  at  $\mathbf{u} = \mathbf{0}$ . Using a computer implementation with an exact rational LP solver QSOpt [1] we found other examples of functions  $f$  with  $n = 4$  variables which have tight bisubmodular relaxations  $g$  (i.e.  $g$  has a minimizer in  $\mathcal{X}^\circ$ ), but all submodular relaxations are not tight. One such example is shown in Fig. 1(d); in this example, the minima of the tightest bisubmodular and submodular relaxations are 0 and  $-3/10$ , respectively.

## References

- [1] Asem M. Ali, Aly A. Farag, Georgy L. Gimel'Farb, Optimizing binary MRFs with higher order cliques, in: ECCV, 2008.
- [2] Kazutoshi Ando, Satoru Fujishige, Takeshi Naitoh, A characterization of bisubmodular functions, Discrete Math. 148 (1996) 299–303.
- [3] M.L. Balinski, Integer programming: methods, uses, computation, Manag. Sci. 12 (3) (1965) 253–313.
- [4] E. Boros, P.L. Hammer, Pseudo-Boolean optimization, Discrete Appl. Math. 123 (1–3) (2002) 155–225.
- [5] E. Boros, P.L. Hammer, X. Sun, Network flows and minimization of quadratic pseudo-Boolean functions, Technical Report RRR 17-1991, RUTCOR, May 1991.
- [6] E. Boros, P.L. Hammer, G. Tavares, Preprocessing of unconstrained quadratic binary optimization, Technical Report RRR 10-2006, RUTCOR, 2006.
- [7] A. Bouchet, Greedy algorithm and symmetric matroids, Math. Program. 38 (1987) 147–159.
- [8] A. Bouchet, W.H. Cunningham, Delta-matroids, jump systems and bisubmodular polyhedra, SIAM J. Discrete Math. 8 (1995) 17–32.
- [9] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, Pattern Anal. Mach. Intell. 23 (11) (2001).
- [10] R. Chandrasekaran, Santosh N. Kabadi, Pseudomatroids, Discrete Math. 71 (1988) 205–217.
- [11] S. Fujishige, Submodular Functions and Optimization, North-Holland, 1991.
- [12] Satoru Fujishige, Satoru Iwata, Bisubmodular function minimization, SIAM J. Discrete Math. 19 (4) (2006) 1065–1073.
- [13] P.L. Hammer, P. Hansen, B. Simeone, Roof duality, complementation and persistency in quadratic 0–1 optimization, Math. Program. 28 (1984) 121–155.
- [14] D. Hochbaum, Instant recognition of half integrality and 2-approximations, in: 3rd International Workshop on Approximation Algorithms for Combinatorial Optimization, 1998.

- [15] D. Hochbaum, Solving integer programs over monotone inequalities in three variables: a framework for half integrality and good approximations, *European J. Oper. Res.* 140 (2) (2002) 291–321.
- [16] H. Ishikawa, Higher-order clique reduction in binary graph cut, in: CVPR, 2009.
- [17] H. Ishikawa, Higher-order gradient descent by fusion-move graph cut, in: ICCV, 2009.
- [18] Satoru Iwata, Kiyohito Nagano, Submodular function minimization under covering constraints, in: FOCS, October 2009.
- [19] Santosh N. Kabadi, R. Chandrasekaran, On totally dual integral systems, *Discrete Appl. Math.* 26 (1990) 87–104.
- [20] V. Kolmogorov, Minimizing a sum of submodular functions, Technical Report, June 2010. [arXiv:1006.1990v1](https://arxiv.org/abs/1006.1990v1).
- [21] V. Kolmogorov, Submodularity on a tree: unifying  $L^2$ -convex and bisubmodular functions, in: 36th International Symposium on Mathematical Foundations of Computer Science, MFCS, August 2011.
- [22] Victor Lempitsky, Carsten Rother, Andrew Blake, LogCut—efficient graph cut optimization for Markov random fields, in: ICCV, 2007.
- [23] Victor Lempitsky, Carsten Rother, Stefan Roth, Andrew Blake, Fusion moves for Markov random field optimization, *Pattern Anal. Mach. Intell.* (2009).
- [24] S.H. Lu, A.C. Williams, Roof duality for polynomial 0–1 optimization, *Math. Program.* 37 (3) (1987) 357–360.
- [25] S. Thomas McCormick, Satoru Fujishige, Strongly polynomial and fully combinatorial algorithms for bisubmodular function minimization, *Math. Program. Ser. A* 122 (2010) 87–120.
- [26] M. Nakamura, A characterization of greedy sets: universal polymatroids (I), in: *Scientific Papers of the College of Arts and Sciences*, vol. 38 (2). The University of Tokyo, 1998, pp. 155–167.
- [27] G.L. Nemhauser, L.E. Trotter, Vertex packings: structural properties and algorithms, *Math. Program.* 8 (1975) 232–248.
- [28] Liqun Qi, Directed submodularity, ditroids and directed submodular flows, *Math. Program.* 42 (1988) 579–599.
- [29] A. Raj, G. Singh, R. Zabih, MRF's for MRI's: Bayesian reconstruction of MR images via graph cuts, in: CVPR, 2006.
- [30] Stefan Roth, Michael J. Black, Fields of experts, *Int. J. Comput. Vis.* 82 (2) (2009) 205–229.
- [31] C. Rother, V. Kolmogorov, V. Lempitsky, M. Szummer, Optimizing binary MRFs via extended roof duality, in: CVPR, June 2007.
- [32] O. Woodford, P. Torr, I. Reid, A. Fitzgibbon, Global stereo reconstruction under second order smoothness priors, in: CVPR, 2008.